Molecular biological methods for studying the gut microbiota: the EU human gut flora project

M. Blaut1*, M. D. Collins2, G. W. Welling3, J. Doré4, J. van Loo5 and W. de Vos6

1German Institute of Human Nutrition, Potsdam-Rehbruecke, Department of Gastrointestinal Microbiology, Arthur-Scheunert-Allee 114–116, 14558 Bergholz-Rehbruecke, Germany
2Food Microbial Sciences Unit, School of Food Biosciences, The University of Reading, PO Box 226, Whiteknights, Reading, RG6 6AP, UK
3Rijksuniversiteit Groningen, Laboratorium voor Medische Microbiologie, Hanzeplein 1, 9713 GZ Groningen, The Netherlands
4Institut National de la Recherche Agronomique, Laboratoire d’Ecologie et de Physiologie du Système Digestif, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France
5ORAFTI, Aandorenstraat 1, 3300 Tienen, Belgium
6Wageningen University, Laboratory of Microbiology, Hesselink van Suchtelenweg 4, Wageningen 6703 CT, The Netherlands

Seven European laboratories co-operated in a joint project (FAIR CT97-3035) to develop, refine and apply molecular methods towards facilitating elucidation of the complex composition of the human intestinal microflora and to devise robust methodologies for monitoring the gut flora in response to diet. An extensive database of 16S rRNA sequences for tracking intestinal bacteria was generated by sequencing the 16S rRNA genes of new faecal isolates and of clones obtained by amplification with polymerase chain reaction (PCR) on faecal DNA from subjects belonging to different age groups. The analyses indicated that the number of different species (diversity) present in the human gut increased with age. The sequence information generated, provided the basis for design of 16S rRNA-directed oligonucleotide probes to specifically detect bacteria at various levels of phylogenetic hierarchy. The probes were tested for their specificity and used in whole-cell and dot-blot hybridisations. The applicability of the developed methods was demonstrated in several studies and the major outcomes are described.

Intestinal flora: Gut microbiology: Ribosomal RNA: Genotypic identification of bacteria

Introduction

Diet has direct effects on the human host such as a supply of nutrients but also indirect effects that are mediated by the intestinal microflora. Micro-organisms resident in the human intestinal tract influence host physiology including gut metabolism, regulation of epithelial cell growth and immune function (Rowland, 1988; Cummings & Macfarlane, 1997; Berg, 1996). Proposed dietary strategies to improve health and well-being of the host include the consumption of pre- and synbiotics. Both are aimed towards stimulation of the growth of beneficial bacteria and suppression of potentially harmful micro-organisms (Roberfroid, 1998). However, the factors controlling the human intestinal flora are largely unknown. In addition, knowledge of the composition of the human gut microbiota is far from complete because, until recently, the identification of bacteria was almost entirely based on phenotypic characteristics. These are often unreliable and lack the resolving power necessary to accurately analyse the extremely complex microbiota of the human gut. Moreover, classical microbiological methodologies of assessing microflora composition are time-consuming, laborious and costly.

Microbial diversity within the human gut is currently grossly underestimated. However, the advent of molecular biology has revolutionised the identification of micro-organisms. In particular the analysis of ribosomal RNA (rRNA) sequences provides an immensely powerful tool for determining the evolutionary interrelationships of...
micro-organisms (Woese, 1987). The high specificity and cumulative nature of rRNA sequence data have spurred the discovery and recognition of new biodiversity.

Because of the shortcomings of phenotypic approaches, seven European laboratories have undertaken a joint effort to develop, refine and apply molecular methods for elucidating the complex structure of the intestinal microflora and to devise methods for monitoring the gut flora in response to diet and the relevance for health. This project (FAIR CT97-3035) was funded by the European Union under the FAIR programme and carried out over a period of 3 years. The construction of a comparative phylogenetic framework of gut micro-organisms based on a highly discriminatory sequence database of 16S rRNA sequences was one of the major tasks in this project with the objective to facilitate recognition of hitherto unknown bacterial species and design of diagnostic probes.

A molecular approach to studying the gut microbiota

Description of new microbial diversity

Ribosomal RNAs are excellent molecules for measuring evolutionary relationships among organisms (Olsen et al. 1986). In contrast to traditional taxonomy, which is based on phenotypic traits, this kind of taxonomy reflects natural evolutionary relationships among organisms (Woese et al. 1990). Prokaryotic ribosomes contain two subunits, the sizes being 50S and 30S (Fig. 1). The 50S subunit contains about thirty-four proteins as well as 5S rRNA (120 bases), and 23S rRNA (about 2900 bases), and the 30S subunit contains about twenty-one proteins and 16S rRNA (about 1500 bases). The 16S rRNA has been the most widely employed molecule to develop the phylogeny of prokaryotes. Within the various regions of 16S rRNA, the degree of conservation differs considerably. Analyses of rRNA sequences have revealed signature sequences, short stretches of rRNA, that are unique to a certain group or groups of organisms enabling the phylogenetic placement and identification of bacteria. By utilising nucleic acids derived directly from natural microbial communities combined with polymerase chain reaction (PCR) and cloning strategies, even non-culturable micro-organisms become accessible to characterisation and identification.

The culture-independent approach used in the project (Fig. 2) comprised random amplification of the 16S rDNA gene on DNA isolated from a faecal sample of a healthy human adult (Suau et al. 1999). Using 10-cycle PCR amplifications and primers binding to conserved sites of the 16S rRNA gene, 284 clones were obtained. These could be classified into eighty-two molecular species (gathering sequences with ≥98% similarity). The majority (95%) of clones could be assigned to the Bacteroides/Prevotella group, the Clostridium coccoides group or the Clostridium leptum subgroup (Fig. 3). The remaining clones were distributed among a variety of phylogenetic clusters. Interestingly, 76% of the characterised 16S rDNA molecular species corresponded to hitherto undescribed bacteria, i.e. only 24% of the molecular species corresponded to

---

**Fig. 1.** The ribosomal RNA as a phylogenetic and evolutionary marker.
Fig. 2. The culture independent assessment of microbial diversity.

Fig. 3. Phylogenetic tree of major phylogenetic groups of intestinal bacteria.
organisms, for which 16S rRNA sequences were available in public databases.

In parallel, 25-cycle PCR was performed on DNA isolated from faeces of the same individual and the resulting amplified 16S rDNA sequenced. The sequence information obtained using either the 10- or 25-cycles approach was compared. A mathematical model to estimate coverage value has been developed and evaluated in order to facilitate comparison of the structure of the two 16S rDNA amplified communities (i.e. 10- v. 25-cycles). After detailed statistical analysis, it was demonstrated that 25-cycle clones were less diversified, in spite of exhibiting a diversity partially different from that of 10-cycle clones (unpublished results). Similar findings were reported by Wilson & Blitchington (1996). Although amplification, cloning and sequencing is a very useful tool to get information on the diversity of a microbial community, it appears that certain bacteria are under-represented in clone libraries of the human gut microbial community. To avoid a biased view of the ecosystem, PCR-independent methods such as in situ hybridisation have to be used (see below).

**Increase in microbial diversity with age**

Direct PCR analyses were also performed on an elderly person’s faecal sample. Over 280 clones were generated and characterised by sequence analyses, providing a molecular taxonomic inventory. Phylogenetic analysis showed that the flora of the elderly was even more diversified than that of the adults (Table 1). The proportion of unknown molecular species was very high among the clones derived from the analysis of an elderly person’s faecal sample. Interestingly, 22% of the elderly person’s clones derived from the analysis of an elderly person’s unknown molecular species was very high among the fied than that of the adults (Table 1). The proportion of showed that the flora of the elderly was even more diversi-

**Development of oligonucleotide probes for the culture-independent identification of human gut bacteria**

The 16S rRNA sequence analyses performed in the European project (FAIR CT97-3035) have greatly advanced our knowledge on the true genetic diversity of the microorganisms in the human gut. Since 16S rRNA contains stretches that are characteristic of different taxa, e.g. from species to generic or suprageneric groups, they can be exploited for the design of characteristic DNA oligonucleotide probes for hybridisation to facilitate the identification of intestinal bacteria at different levels of the taxonomic hierarchy. A variety of probing strategies has been developed, such as microscopy-based fluorescent in situ hybridisation or quantitative dot-blot hybridisation (Amann et al. 1990b; Stahl et al. 1988). These probing technologies have been extensively applied to some natural ecosystems, but until recently, their use in gut microbiology has been very limited.

Owing to the great potential of such technologies, one of the major objectives in this project was the design of oligonucleotide probes targeting the 16S rRNA of intestinal micro-organisms. Probe design necessitates a comparison of the 16S rRNA sequence of the target organism, with rRNA sequences already in the database, to find a specific region. Software packages such as the ARB program greatly facilitate probe design (Ludwig et al. 1998). Any probe designed is subsequently checked using the Check-probe function of the Ribosomal Database Project software package (Maidak et al. 1997). Based on a highly improved database of 16S rRNA sequences, new oligonucleotide probes targeting intestinal bacteria were designed and, together with probes already available from work of other researchers, validated, i.e. the specificity of these probes was tested with up to 100 bacterial species representing dominant groups of intestinal bacteria. The probes target bacteria at various levels of the phylogenetic hierarchy. Oligonucleotides targeting all living cell or cells at the domain level (Eucarya, Archaea, or Bacteria) have been designed previously (Zheng et al. 1996; Amann et al. 1990a) and a few probes detecting numerically pre-

**Table 1. Increase in microbial diversity with age**

<table>
<thead>
<tr>
<th>Subjects (number)</th>
<th>Number of clones</th>
<th>Number of species</th>
<th>% described species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infants (2)</td>
<td>164</td>
<td>15</td>
<td>70</td>
</tr>
<tr>
<td>Adults (5)</td>
<td>619</td>
<td>160</td>
<td>19</td>
</tr>
<tr>
<td>Elderly (1)</td>
<td>280</td>
<td>168</td>
<td>8</td>
</tr>
</tbody>
</table>
The first probes designed and validated in the project were aimed to detect intestinal bacteria at the group level. These included probes for the \textit{Bifidobacterium} group, Gram-positive bacteria with a low GC (guanosine and cytosine) content, the \textit{Clostridium histolyticum} group, the \textit{Clostridium lituseburense} group, the \textit{Eubacterium rectale/Clostridium coccoides} group, the \textit{Bacteroides fragilis} and the \textit{Bacteroides distasonis} subgroups, the \textit{Clostridium leptum} subgroup, and the enteric bacteria. (Wilson & Blitchington, 1996; Langendijk \textit{et al.} 1995; Franks \textit{et al.} 1998; Sghir \textit{et al.} 2000). The phylogenetic relationship of these groups is shown in Fig. 3. Subsequently, oligonucleotide probes targeting subgroups or even species were designed and applied (Harmsen \textit{et al.} 2000; Dore` \textit{et al.} 1998; Simmering \textit{et al.} 1999; Schwiertz \textit{et al.} 2000).

There are two different ways to use 16S rRNA-targeted oligonucleotide probes for quantifying bacterial populations: quantitative dot-blot hybridisation and whole-cell hybridisation. Neither method depends upon culturing. With the former, the amount of 16S rRNA of a target organism is related to the amount of 16S rRNA of all organisms present in a given sample. In practice, this is done by isolating ribosomal RNA from a faecal sample and blotting known amounts of it, in parallel, onto two nylon membranes (Fig. 4). One membrane is subsequently hybridised with a universal oligonucleotide probe that detects all micro-organisms because it recognises a conserved target region in the 16S rRNA molecule. The other membrane is hybridised with the specific probe designed for detection of the target organism(s). The probes are usually labelled with $^{32}$P to be able to quantify rRNA sequences with a relatively low abundance of 0.1 – 1%. The relative abundance of the target organism is calculated by dividing the signal obtained with the specific probe, at a given rRNA concentration, by the signal obtained with the universal probe or with the fluorescent dye 4',6-diamidino-2-phenylindol (DAPI). It should be noted that these numbers cannot be directly translated into cell numbers because cells of different species have different ribosome contents ranging from $10^3$ to $10^5$ ribosomes per cell and the ribosome content also varies with growth rate (Amann \textit{et al.} 1995).

Whole-cell hybridisation combined with specific 16S rRNA-targeted oligonucleotide probes detects morphologically intact bacterial cells (Fig. 5). Since fluorescently labelled probes are usually employed, this technique is often referred to as fluorescence \textit{in situ} hybridisation (FISH). In principal, bacterial cells are permeabilised by treatment with paraformaldehyde or ethanol to allow the probes to reach their target, the ribosomal RNA. If the 16S rRNA contains a sequence complementary to the probe sequence and if this target sequence is accessible to the fluorescent oligonucleotide, a hybrid is formed causing the whole cell to fluoresce. Since each cell contains several thousand ribosomes, the fluorescing cells can be visualised by epifluorescence microscopy. The major advantage of this method is the capability to detect individual cells in a complex mixture of cells without the need to grow them.

**Application of 16S rRNA-targeted oligonucleotide probes in human faecal flora studies**

Both methods have been applied to human faecal samples to determine the composition of the bacterial flora. The number of probes employed in these studies was limited and primarily took advantage of the group-level probes developed. Franks and collaborators investigated the variations of predominant bacterial populations in faeces of nine volunteers over a period of 8 months using whole.
cell *in situ* hybridisation (Franks *et al.* 1998). Two probes specific for the *Bacteroides fragilis* and the *Bacteroides distasonis* subgroups detected 20% of the faecal flora while a probe specific for the *Eubacterium rectale/Clostridium cocoides* group detected 29%. The other three probes used for the identification of the *Clostridium lituseburenses* group, the *Clostridium histolyticum* group and the *Streptococcus–Lactococcus* group, detected less than 1% of the faecal flora. Nevertheless, these probes were considered important because pathogens related to certain gut disorders may be detected (Franks *et al.* 1998). On average, the proportion of the bifidobacteria was 3% of the faecal flora. This result is in agreement with that of an earlier study in which it had also been shown that the proportion of bifidobacteria is overestimated when cultural methods are used (Langendijk *et al.* 1995). This may be explained by the finding that total bacterial counts determined through cultural methods are considerably smaller than microscopic counts. The probe specific for Gram-positive bacteria with a low GC content detected 12% of the faecal bacterial cells (Franks *et al.* 1998). Taken together, all probes used in this study accounted for approximately two thirds of the normal human faecal flora. The study also revealed that the faecal flora underwent fluctuations over time, but that differences in flora composition between volunteers and over time can be assessed objectively.

A similar study, but based on quantitative dot-blot hybridisation, was performed by Sghir *et al.* (2000) to investigate the population structure of the predominant phylogenetic groups within the human adult faecal microbiota. These researchers applied six probes to faecal samples collected from twenty-seven human adults. The probes accounted for 70% of the total 16S rRNA detected by the bacterial domain probe. The probes specific for the *Bacteroides* group, the *Clostridium leptum* group, the *Clostridium cocoides/Eubacterium* group, the *Bifidobacterium* group, the *Lactobacillus* group and the enterobacterial group accounted for 37%, 16%, 14%, 0.7%, 0.6%, and 0.7%, respectively, of the counts detected by the bacterial domain probe. This study revealed variations between individuals but no striking differences between males and females. The two studies (Sghir *et al.* 2000; Franks *et al.* 1998) cannot be directly compared because the probes used were not identical and, as already pointed out, the methods used for quantification differed. In the study using quantitative dot-blot hybridisation (Sghir *et al.* 2000), the probe targeting the *Bacteroides*, *Prevotella* and *Porphyromonas* group was broader than that used in the FISH-based study (Franks *et al.* 1998), in which a combination of probes targeting *Bacteroides fragilis* and *Bacteroides distasonis* were used, accounting for 37% and 20% of the faecal flora, respectively. In spite of the fact that both studies employed the same probe for the detection of the *Clostridium cocoides/Eubacterium rectale* cluster, the relative proportion of this phylogenetic population group differed considerably; 16% (Sghir *et al.* 2000) v. 29% (Franks *et al.* 1998). This clearly demonstrates that relative proportions of cell numbers on the one hand and of ribosomal RNA on the other hand do not necessarily...
agree. In addition, it has to be kept in mind that the individuals investigated in the two studies were not identical.

Oligonucleotide probes were also used in a controlled feeding trial to monitor faecal bifidobacteria in response to the oral intake of inulin (Krus~e et al. 1999). Eight healthy subjects consumed a fat-reduced diet that provided 30% energy as fat and 55% energy as carbohydrates for a period of 64 days using inulin as a fat replacer. The amounts of inulin consumed were based on individual energy requirements such that subjects consumed up to 34 g per day. Inulin led to a significant increase in bifidobacteria by more than one order of magnitude. The data collected in this study revealed high individual variability between subjects. The results corroborate previous in vivo studies in which classical microbiological techniques were used for the enumeration of bifidobacteria (Gibson & Roberfroid, 1995; Bouhnik et al. 1996).

In the project, efforts were also made to quantify human faecal bacteria that are difficult to identify by classical microbiological techniques. Therefore, oligonucleotide probes were developed for human intestinal species of the genus Eubacterium (Schwietz et al. 2000). In the human intestinal tract, Eubacterium is thought to be the second most common genus after Bacteroides (Finegold et al. 1983). However, the identification of Eubacterium species based on phenotypic characteristics is both difficult and time-consuming. Moreover, the taxon Eubacterium is phylogenetically diverse, preventing the design of a Eubacterium group-specific probe. In order to quantify different Eubacterium species in human faecal samples, species-specific 16S rRNA-targeted oligonucleotide probes were designed and validated. The choice of bacteria was based on studies by Finegold et al. (1983) and Moore & Holdeman (1974). Probes were developed for Eubacterium barkeri, E. biforme, E. contortum, E. cylindroides, E. dolichum, E. hadrum, E. lentum, E. limosum, E. moniliforme and E. ventriosum and subsequently applied to human faecal samples obtained from twelve healthy adults. E. barkeri, E. dolichum, E. limosum, and E. moniliforme were not detected in any of the twelve individuals analysed, whereas the other Eubacterium species mentioned previously were found to be present in at least one subject at cell numbers that correspond to 0.03%–1% of the total bacterial cells. All Eubacterium probe-positive infants hybridised also with the Coriobacterium probe indicating that they all belonged to this group. The proportion of bifidobacteria was smaller in formula-fed infants than in breast-fed infants. In contrast to the situation in babies, in which all cells recognised with the Atopobium probe belong exclusively to the Coriobacterium group, older children or adults harbour in addition Eggerthella or Atopobium (Harmsen et al. 2000).

Oligonucleotide probes have also been used to track a single species in the faeces of human subjects or experimental animals in order to investigate numerical importance. A good example is Eubacterium ramulus, an organism that was isolated from human faeces and shown to be able to anaerobically degrade the flavonoid quercetin-3-glucoside to 3,4-dihydroxyphenylacetic acid (Fig. 6) (Schneider et al. 1999). Quercetin is a component found in many plants that are relevant to human nutrition. It is commonly linked to sugar moieties and has received much attention in recent years due to its antioxidative and possibly anticarcinogenic activities (Suolinnna et al. 1975). Gnotobiotic rats associated with Eubacterium ramulus, but not germ-free rats, excrete 3,4-dihydroxyphenylacetic acid in their urine and faeces in response to the oral application of quercetin-3-glucoside. This indicates that the bacterium, whose concentration in intestinal contents and faeces was monitored by FISH, catalyses the same metabolic transformation observed previously under in vitro conditions (Schneider et al. 2000).

Recently, 16S rRNA-based probes targeting the Coriobacterium group and the Atopobium cluster were developed and applied to faecal samples of breast-fed and formula-fed infants and of adults from three different age groups (Harmsen et al. 2000). The Coriobacteriaceae comprise anaerobic bacteria that can be cultured readily from human faeces (Holdeman et al. 1976). Examples are Eggerthella lenta (formerly Eubacterium lentum) and Collinsella aerofaciens (formerly Eubacterium aerofaciens) both of which are members of the human intestinal flora (Finegold et al. 1983). The Coriobacteriaceae belong to the Actinobacteria subdivision, which comprises in addition the genus Atopobium. One of the newly developed probes used in this study was specific for the Atopobium cluster, which includes Eggerthella lenta, Collinsella aerofaciens, and the genera Coriobacterium and Atopobium. Another probe covered only a subgroup of this cluster, namely Collinsella aerofaciens and the genus Coriobacterium. Since bifidobacteria are known to be a numerically dominant group in the intestine of infants a Bifidobacterium-specific probe was also included in the study (Harmsen et al. 2000). These probes were applied to faecal samples from breast-fed and formula-fed infants and analysed by FISH. In breast-fed babies, one out of six babies contained small numbers of Coriobacterium-group cells, while in formula-fed babies large numbers of these bacteria were found in their faeces, reaching up to 39% of the total bacterial cells. All Atopobium probe-positive infants hybridised also with the Coriobacterium probe indicating that they all belonged to this group. The proportion of bifidobacteria was smaller in formula-fed infants than in breast-fed infants.

![Fig. 6. Conversion of quercetin-3-glucoside as catalysed by faecal Eubacterium ramulus.](https://doi.org/10.1079/BJN/2002539)
the occurrence of *E. ramulus* in the human intestinal tract, an oligonucleotide probe specific for this organism was designed and validated (Simmering et al. 1999). Application of the probe to faecal samples from twenty subjects aged 23 – 59 years indicated the presence of *E. ramulus* in each of the individuals tested at concentrations of $4 \times 10^7$ to $2 \times 10^9$ cells/g of faecal dry mass. Interestingly, six different faecal isolates identified as *E. ramulus* exhibited different banding patterns when analysed by Randomly Amplified Polymorphic DNA (RAPD) profiles, indicating that they were different strains of the same species.

### Conclusions

The application of molecular tools to intestinal microbiology has greatly facilitated study of the complex microbial community resident in the human intestinal tract. The number of new 16S rRNA sequences determined in the course of the European project FAIR CT97-3035 has greatly improved the database on gut microbes but also provided a good basis for a concept of the diversity of intestinal micro-organisms. Data gained in this project clearly show that there is an increase in the number of bacterial species with age. Owing to the highly improved database it was possible to design and validate a considerable number of probes that target intestinal bacteria at different levels of hierarchy. However, the diversity of the human intestinal flora turned out to be much greater than anticipated at the beginning of the project. A considerable number of oligonucleotide probes have become available for the identification of gut micro-organisms. They provide an excellent basis for reliable microflora analyses. The usefulness of FISH and dot-blot hybridisation, all of which take advantage of oligonucleotide probes, has been demonstrated in several independent studies performed during the course of the project. Based on the work completed in this project, monitoring of the faecal flora in response to dietary intervention has become feasible.

Future activities will be directed at making this molecular approach easy to use and applicable to the analysis of large numbers of human faecal samples. To improve resolution and to extend the number of diagnostic probes, the comparative 16S rRNA database for the human intestinal microflora has to become more fully comprehensive. Limitations in the number of samples that can be analysed have to be overcome by 16S rRNA-based high throughput approaches that afford the automatic detection of either fluorescently labelled cells or of isolated rRNA species hybridised to DNA arrays.

Knowledge of the identity of bacteria in the human gut ecosystem is not sufficient to understand microbe/microbe and microbe/host interactions in the gut. Therefore, an *in situ* approach has to be developed to assess selected microbial activities in individual cells. The *in situ* detection of mRNA from bacteria or the use of fluorogenic substrates should be further developed and applied to human gut microbiota. Finally, it has to be emphasised that classical microbiological techniques are not obsolete but complementary to the molecular techniques being developed.

### Acknowledgements

The work summarised in this paper has been carried out with financial support from the Commission of the European Communities (FAIR-CT97-3035). It does not necessarily reflect its views nor in anyway anticipates the Commission’s future policy in this area. The authors also would like to acknowledge the valuable work of their coworkers whose contribution was essential to make this project successful.

### References


Holdeman LV, Good IJ & Moore WEC (1976) *Human fecal flora: variation in bacterial composition within individuals* and a

* Note: The references marked with an asterisk describe work summarised in this paper.


